

# Frugal Online Incentive Mechanisms for Crowdsourcing Tasks Truthfully

Dong Zhao, Huadong Ma, *Member, IEEE*, and Liang Liu

**Abstract**—Mobile Crowd Sensing (MCS) is a new paradigm which takes advantage of pervasive smartphones to efficiently collect data, enabling numerous novel applications. To achieve good service quality for a MCS application, incentive mechanisms are necessary to attract more user participation. Most of existing mechanisms apply only for the *offline* scenario where all users' information are known a priori. On the contrary, we focus on a more realistic scenario where users arrive one by one *online* in a random order. Based on the *online auction* model, we investigate the problem that users submit their private profiles to the crowdsourcer when they arrive, and the crowdsourcer aims at selecting a subset of users before a specified deadline for minimizing the total payment while a specific number of tasks can be completed. We design three *online mechanisms*, *Homo-OMZ*, *Hetero-OMZ* and *Hetero-OMG*, all of which can satisfy the *computational efficiency*, *individual rationality*, *cost-truthfulness*, and *consumer sovereignty*. The *Homo-OMZ* mechanism is applicable to the homogeneous user model and can satisfy the *social efficiency* but not *constant frugality*. The *Hetero-OMZ* and *Hetero-OMG* mechanisms are applicable to both the homogeneous and heterogeneous user models, and can satisfy the *constant frugality*. Besides, the *Hetero-OMG* mechanism can also satisfy the *time-truthfulness*. Through extensive simulations, we evaluate the performance and validate the theoretical properties of our online mechanisms.

**Index Terms**—Crowdsourcing, Mobile Sensing, Incentive Mechanism, Frugal Mechanism, Online Auction.

## I. INTRODUCTION

Crowdsourcing is a distributed problem-solving model in which a crowd of undefined size is engaged to solve a complex problem through an open call [1]. Nowadays, the proliferation of smartphones provides a new opportunity for extending existing web-based crowdsourcing applications to a larger contributing crowd, making contribution easier and omnipresent. Furthermore, today's smartphones are programmable and come with a rich set of cheap powerful embedded sensors, such as GPS, WiFi/3G/4G interfaces, accelerometer, digital compass, gyroscope, microphone, and camera. The great potential of mobile phone sensing offers a variety of novel, efficient ways to opportunistically collect data, enabling numerous *mobile crowd sensing* (MCS) applications, such as Sensorly [2] for constructing cellular/WiFi network coverage maps, SignalGuru [3], Nerice [4] and VTrack [5] for providing traffic information, Ear-Phone [6] and NoiseTube [7] for making noise maps,

and LiFS [8] for indoor localization. For more details, we refer interested readers to several survey papers [1], [9], [10].

Adequate user participation is one of the most critical factors determining whether a MCS application can achieve good service quality. Most of the current MCS applications [2]–[8] are based on voluntary participation. While participating in a MCS campaign, smartphone users consume their own resources such as battery and computing power, and expose their locations with potential privacy threats. Thus, incentive mechanisms are necessary to provide participants with enough rewards for their participation costs. Most of existing mechanisms [11]–[15] apply only for the *offline* scenario as illustrated in Fig. 1(a), in which all of participating users report their profiles, including the tasks that they can complete and the bids, to the crowdsourcer (campaign organizer) in advance, and then the crowdsourcer selects a subset of users after collecting the information of all users to achieve a specific objective. Generally, there are two classes of incentive mechanisms with different objectives: the *budget feasible* mechanisms which aim at maximizing the crowdsourcer's utility (e.g., the total value of all tasks that can be completed by selected users) under a specific budget constraint, and the *frugal* mechanisms which aim at minimizing the crowdsourcer's total payment under the condition that the specific tasks can be completed.

However, in most of MCS applications users always arrive one by one *online* in a random order and user availability changes over time, which reflects the opportunistic characteristics of human mobility in essence. For example, in Sensorly [2] or Ear-Phone [6], the crowdsourcer allocates tasks to the smartphone users for sensing Wi-Fi signals or environmental noises only when the users arrive in the area of interest opportunistically. Therefore, an *online incentive mechanism* is necessary to make irrevocable decisions on whether to accept a user's task and bid *sequentially*, based solely on the information of users arriving before the present moment, without knowing future information, as illustrated in Fig. 1(b) and Fig. 1(c).

Although the *budget feasible* online incentive mechanisms have been investigated recently [16]–[19], in this paper we focus on the online incentive mechanisms under the framework of *frugality*. Specially, we consider the following problem: the crowdsourcer aims at selecting a subset of users before a specified deadline, so that the total payment to the selected users can be minimized under the condition that the specific number of tasks can be completed. We further assume that the cost and arrival/departure

D. Zhao, H.-D. Ma, and L. Liu are with the Beijing Key Lab of Intelligent Telecomm. Software and Multimedia, Beijing University of Posts and Telecomm., Beijing, 100876, China. E-mail: {dzhao, mhd}@bupt.edu.cn, liangliu.bupt@gmail.com.

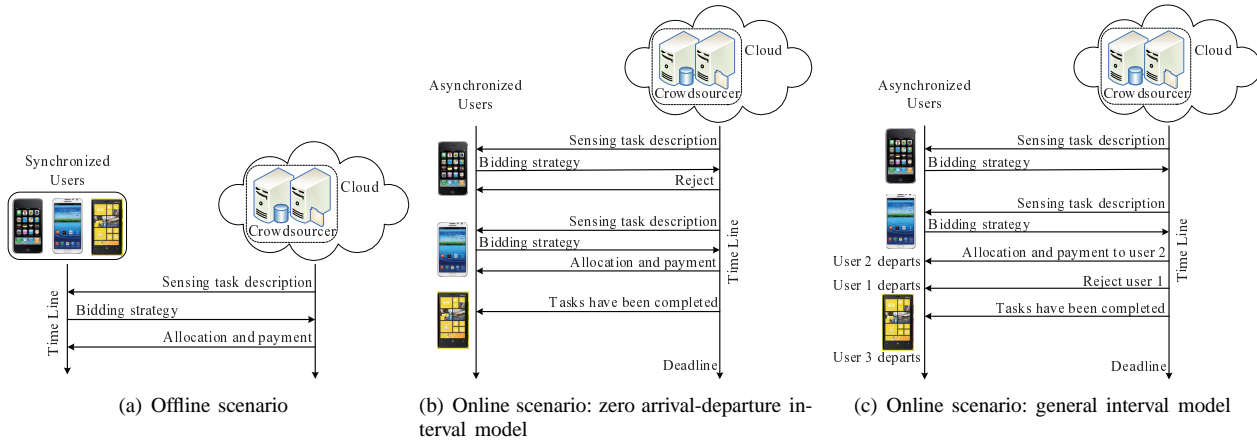


Fig. 1. The comparison of offline and online scenarios for MCS.

TABLE I  
COMPARISON OF VARIOUS MECHANISMS PROPOSED IN THIS PAPER.

Mechanism	homogeneous user model	heterogeneous user model	i.i.d. model	secretary model	computational efficiency	individual rationality	cost-truthfulness	consumer sovereignty	constant frugality	time-truthfulness (general interval model)
Homo-OMZ	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	NA
Hetero-OMZ	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	NA
Hetero-OMG	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

time of each user are private and only known to itself. Users are assumed to be game-theoretic and seek to make strategy (possibly report an untruthful cost or arrival/departure time) to maximize their individual utility in equilibrium. Thus, the problem of selecting crowdsourcing users while minimizing the total payment can be modeled as an *online auction*.

Our objective is to design online mechanisms satisfying five desirable properties: *computational efficiency*, *individual rationality*, *truthfulness*, *consumer sovereignty* and *constant frugality*. Informally, *computational efficiency* ensures the mechanism can run in real time, *individual rationality* ensures each participating user has a non-negative utility, *truthfulness* ensures the participating users report their true costs (*cost-truthfulness*) and arrival/departure times (*time-truthfulness*), *consumer sovereignty* ensures each participating user has a chance to win the auction, and *constant frugality* guarantees that the mechanism performs close to the optimal solution in terms of total payment in the offline scenario where all users' information are known a priori.

In this paper, we consider various user models comprehensively from three perspectives: 1) the *zero arrival-departure interval* model (as illustrated in Fig. 1(b)) and the *general interval* model (as illustrated in Fig. 1(c)) with respect to the interval between the arrival time and the departure time of each user, 2) the *homogeneous* user model and the *heterogeneous* user model with respect to the number of tasks that each user can complete, and 3) the *i.i.d.* model and the *secretary* model with respect to the distribution of users. We propose three online mechanisms, *Homo-OMZ*, *Hetero-OMZ* and *Hetero-OMG*, that are applicable to different models and hold different properties, as shown in Table I. Note that achieving time-truthfulness is trivial in the *zero arrival-departure interval* model. The main idea behind our online mechanisms is to adopt a multiple-stage sampling-accepting process. At every stage the mechanism allocates tasks to a user only if its bid is

not less than a certain bid threshold that has been computed using previous users' information, and the number of tasks allocated for the current stage has not been achieved. Meanwhile, the user obtains a bid-independent payment. The bid threshold is computed in a manner that guarantees desirable performance properties of the mechanism.

The remainder of this paper is organized as follows. In Section II we describe the MCS system model, and formulate the problem as an online auction. We present two online mechanisms, *Homo-OMZ* and *Hetero-OMZ*, under the *zero arrival-departure interval* model in Section III. We present the *Hetero-OMZ* mechanism under the *general interval* model in Section IV. Performance evaluations are presented in Section V. We review the related work in Section VI, and conclude this paper in Section VII.

## II. SYSTEM MODEL AND PROBLEM FORMULATION

### A. System Model

As illustrated in Fig. 1, a MCS system consists of a *crowdsourcer*, which resides in the cloud and consists of multiple sensing servers, and many smartphone *users*, which are connected to the cloud by cellular networks (e.g., GSM/3G/4G) or Wi-Fi connections. The crowdsourcer first publicizes a MCS campaign, aiming at finding some users to complete a specified number  $L \in \mathbb{N}^+$  of tasks. Then the users interested in participating in the campaign report their profiles to the crowdsourcer. Finally, the crowdsourcer selects a subset of users to perform tasks for minimizing the total payment based on the profiles of users. Fig. 1(a) shows an offline scenario, where all of participating users report their profiles to the crowdsourcer synchronously, and then the crowdsourcer allocates tasks to a subset of users by considering the profiles of all users at once. Unlike the batched and synchronized manner in the offline scenario, the interactive process in the online scenario is

sequential and asynchronous, as shown in Fig. 1(b) and Fig. 1(c). In this paper we focus on the online scenario. Specially, we assume that the crowdsourcer requires  $L$  tasks to be completed before a specific deadline  $T$ , and a crowd of users  $\mathcal{U} = \{1, 2, \dots, n\}$  interested in participating in the campaign arrive online in a random order, where  $n$  is unknown. Each user  $i \in \mathcal{U}$  has an arrival time  $a_i \in \{1, \dots, T\}$ , a departure time  $d_i \in \{1, \dots, T\}$ ,  $d_i \geq a_i$ , and a number of tasks it can complete  $\tau_i$ . Meanwhile, user  $i$  also has an associated cost  $c_i \in \mathbb{R}^+$  for performing a single task. All information constitutes the *profile* of user  $i$ ,  $\theta_i = (a_i, d_i, \tau_i, c_i)$ . The crowdsourcer, when presented with the profile of user  $i$ , must decide how many tasks to allocate, and how much payment to pay to user  $i$  before it departs. The decisions should be made one by one until the deadline or all the tasks have been allocated.

### B. User Model

In this paper we consider two models with respect to the interval between the arrival time and the departure time of each user:

- **Zero arrival-departure interval model:** The arrival time of each user equals to its departure time.
- **General interval model:** There are no limitations on the arrival-departure interval (i.e., it may be zero or non-zero interval).

The first model is reasonable for the MCS applications where the decision has to be made in time, as illustrated in Fig. 1(b). For example, in LiFS [8], the users receive task description when they enter the target building, and report the profiles expecting the platform to reply immediately, since they may not want to be disturbed anymore when they are working or shopping in that building. Nevertheless, in other scenarios, smartphone users may be patient to wait for the reply of the crowdsourcer for some time interval. For example, a user who is staying in a traffic tool or drinking at a coffee shop may play with the crowdsourcer for some time. In such scenarios, the general interval model should be considered where the decision could be put off until when the user departs, as illustrated in Fig. 1(c).

With respect to the number of tasks that each user can complete, we consider the following two models:

- **Homogeneous user model:** Each user can complete only a single task.
- **Heterogeneous user model:** Different users can complete different number of tasks.

Obviously, the homogeneous user model is a special case of the heterogeneous user model, and in the first model, we denote the *profile* of each user  $i$  as  $\theta_i = (a_i, d_i, 1, c_i)$ .

With respect to the distribution of users, there are two extremes of the modeling spectrums. At one extreme is the *oblivious adversarial model*, where an adversary chooses a *worst-case* input sequence including the users' costs, task numbers, and their arrival orders. This view leads to a very pessimistic result: no online mechanisms can achieve constant competitiveness [20], but this model is unnecessary to consider in a normal crowdsourcing market. At the

other extreme is a sequence of users with task numbers and the respective unit costs i.i.d. sampled from a *known* distribution. This view may lead to the optimum strategy by using dynamic programming, but it assumes too much knowledge to be practical. Thus, two natural candidate models in between the two extremes are considered in this paper:

- **The i.i.d. model:** The numbers of tasks that users can complete and their respective unit costs are i.i.d. sampled from some *unknown* distributions.
- **The secretary model:** An adversary gets to decide on the numbers of tasks that users can complete and the respective unit costs, but not on the *order* in which they are presented to the crowdsourcer.

In fact, the i.i.d. model is a special case of the secretary model, since the sequence can be determined by first picking a multi-set of task numbers and costs from the (unknown) distribution, and then permuting them randomly.

### C. Online Auction

We model the interactive process between the crowdsourcer and users as an *online auction*. Users are assumed to be game-theoretic and seek to make *strategy* to maximize their individual utility in equilibrium. Each user  $i$  makes a *reserve price*  $b_i$ , called *bid*, based on its valuations on each single task. Note that the arrival/departure time and cost of user  $i$  are private and only known to itself. Only the number of tasks  $\tau_i$  must be true since the crowdsourcer can identify whether the announced tasks are performed. In other words, user  $i$  may misreport all information about its profile except for  $\tau_i$ . The crowdsourcer's objective and the deadline  $T$  are common knowledge. Although we do not require a user to declare its departure time until the moment of its departure, we find it convenient to analyze our auctions as direct-revelation mechanisms (DRMs). The strategy space in an online DRM allows a user to declare some possibly untruthful profile  $\hat{\theta}_i = (\hat{a}_i, \hat{d}_i, \tau_i, b_i)$ , subject to  $a_i \leq \hat{a}_i \leq \hat{d}_i \leq d_i$ . Note that we assume that a user cannot announce an earlier arrival time or a later departure time than its true arrival/departure time. In order to complete the required sensing tasks, the crowdsourcer needs to design an *online mechanism*  $\mathcal{M} = (f, p)$  consisting of an *allocation* function  $f$  and a *payment* function  $p$ . For any *strategy sequence*  $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_n)$ , the allocation function  $f(\hat{\theta})$  computes an allocation of tasks for a selected subset of users  $\mathcal{S} \subseteq \mathcal{U}$ , and the payment function  $p(\hat{\theta})$  returns a vector  $(p_1(\hat{\theta}), \dots, p_n(\hat{\theta}))$  of payments per task to the users. That is, each selected user  $i \in \mathcal{S}$  is allocated  $f_i$  tasks at price  $p_i$  per task. Thus, in the homogeneous user model, the *utility* of user  $i$  is

$$u_i = \begin{cases} p_i - c_i, & \text{if } i \in \mathcal{S}; \\ 0, & \text{otherwise.} \end{cases}$$

In the heterogeneous user model, the *utility* of user  $i$  is

$$u_i = \begin{cases} f_i(p_i - c_i), & \text{if } i \in \mathcal{S}; \\ 0, & \text{otherwise.} \end{cases}$$

TABLE II  
FREQUENTLY USED NOTATIONS.

Notation	Description
$\mathcal{U}, n, i$	set of users, number of users, and one user
$L, L'$	total number of tasks and stage-task-number
$T, T', t$	deadline, end time step of each stage, and each time step
$a_i, \hat{a}_i$	true arrival time and strategic arrival time of user $i$
$d_i, \hat{d}_i$	true departure time and strategic departure time of user $i$
$\tau_i$	the number of tasks that user $i$ can complete
$c_i, b_i$	true cost and bid of user $i$ for performing a single task
$\theta_i, \hat{\theta}_i$	true profile and strategy of user $i$
$\mathcal{S}, \mathcal{S}'$	set of selected users and sample set
$f_i, p_i, u_i$	allocation, payment per task, and utility of user $i$
$b^*$	bid threshold
$\delta$	parameter used for computing the bid threshold
$\beta$	initialized bid threshold

Note that, the crowdsourcer, when presented with the strategy  $\hat{\theta}_i$  of user  $i$ , must decide whether to accept user  $i$  at what price ( $p_i$ ) before the time step  $\hat{d}_i$ . In the homogeneous user model, the crowdsourcer expects that the total payment is minimized under the condition that at least  $L$  users are selected before the deadline, i.e.,

$$\text{Minimize } \sum_{i \in \mathcal{S}} p_i \text{ subject to } |\mathcal{S}| = L.$$

In the heterogeneous user model, the crowdsourcer expects that the total payment is minimized under the condition that at least  $L$  tasks are completed before the deadline, i.e.,

$$\text{Minimize } \sum_{i \in \mathcal{S}} f_i p_i \text{ subject to } \sum_{i \in \mathcal{S}} f_i = L.$$

Table II lists frequently used notations in this paper.

#### D. Design Objective

Our objective is to design online mechanisms satisfying the following five desirable properties:

- **Computational Efficiency:** A mechanism is *computationally efficient* if both the allocation and payment can be computed in polynomial time as each user arrives.
- **Individual Rationality:** Each participating user will have a non-negative utility:  $u_i \geq 0$ .
- **Truthfulness:** A mechanism is *cost-* and *time-truthful* (or simply called *truthful*, or *incentive compatible* or *strategyproof*) if reporting the true cost and arrival/departure time is a *dominant strategy* for all users. In other words, no user can improve its utility by submitting a false cost or arrival/departure time, no matter what others submit.
- **Consumer Sovereignty:** The mechanism cannot arbitrarily exclude a user; each user will have a chance to win the auction and obtain a payment if only its bid is sufficiently low while others are fixed.
- **Constant Frugality:** The ideal goal of the mechanism is to minimize the total payment of the crowdsourcer.

Although it is impossible to obtain an optimal solution in the online scenario, we hope that the mechanism can use as small payment as possible for performing all tasks, namely achieving the *frugality*. Ideally, we would like that the total payment made by the mechanism has a constant approximate ratio compared to the minimal cost required for performing all tasks in the offline scenario where the crowdsourcer has full knowledge about users' profiles. We call this ideal objective as "*idealistic frugality*". However, it is easy to know that no truthful mechanism can perform well even if in the offline scenario\*. Therefore, in this paper we use an alternative objective, called as "*realistic frugality*". Specially, we use the *frugality ratio* to measure the frugality as follows: given a number of tasks to perform,  $L$ , we say that a mechanism is  $\alpha$ -*frugal* if it allocates  $L$  tasks in expectation while guaranteeing that the total payment is no more than the minimum cost required to complete  $\alpha L$  tasks in the offline scenario. Here, our goal is to design mechanisms with a constant frugality ratio, namely that the mechanisms can satisfy the *constant frugality*.

The importance of the first two properties is obvious, because they together guarantee that the mechanisms can be implemented in real time and satisfy the basic requirements. In addition, the last three properties are indispensable for guaranteeing the high performance and robustness. The *truthfulness* aims at eliminating the fear of market manipulation and the overhead of strategizing over others for the participating users. The *consumer sovereignty* aims at guaranteeing that each participating user has a chance to win the auction, otherwise it will hinder the users' competition or even result in task starvation. Besides, if some users are guaranteed not to win the auction, then being truthful or not will have the same outcome. For this reason, the property satisfying both the *consumer sovereignty* and *truthfulness* is also called *strong truthfulness* [21]. Later we will show that satisfying the *consumer sovereignty* is not trivial in the online scenario, which is in contrast to the offline scenario. Finally, we emphasize that the *frugality* is different from *social efficiency* which aims at minimizing the *total cost* (not *total payment*) of selected users. It is generally defined that, an online mechanism is  $O(g(n))$ -*competitive* for *social efficiency* if the ratio between the total cost achieved by this mechanism and the optimal offline solution is  $O(g(n))$ .

### III. ONLINE MECHANISM UNDER ZERO ARRIVAL-DEPARTURE INTERVAL MODEL

In this section, we consider the zero arrival-departure interval model where each user is impatient since the decision must be made immediately once it arrives. Note that achieving time-truthfulness is trivial in this model.

\*Consider an instance where a single user can complete all required tasks at a very low cost  $\epsilon$ , while every other user has a very high cost  $C$ . It is not hard to know that any truthful mechanism that completes all tasks must pay at least  $C$ , thus making the ratio between the total payment and the minimal cost unbounded.

It is because that any user has no incentive to report a later arrival time or an earlier departure time than its true arrival/departure time, since the user cannot perform any sensing task or obtain a payment after it departs. To facilitate understanding, it is also assumed that no two users have the same arrival time. Note that this assumption can be easily removed according to the revised mechanism in Section IV.

#### A. Homogeneous User Mechanism

An online mechanism needs to overcome several non-trivial challenges: first, the users' costs are unknown and need to be reported in a truthful manner; second, a specified number of tasks should be completed before the deadline; finally, the mechanism needs to cope with the online arrival of users. In the homogeneous user model, if we do not consider the truthfulness, then the mechanism design problem becomes a  $k$ -choice secretary problem in essence. There are two existing solutions [22], [23] that can achieve a constant-factor approximation, and can be applied into on-line auctions that satisfy the truthfulness. However, none of two solutions can guarantee the *consumer sovereignty*. The first solution [22] adopts a two-stage sampling-accepting process: the first batch of users is rejected and used as the sample which enables making an informed decision on whether accepting the rest of users. In this case, the first batch of users has no chance to win the auction no matter how low its cost is. It can lead to undesirable effects in our problem: automatically rejecting the first batch of users encourages users to arrive late; in other words, those users arriving early have no incentive to report their bids, which may hinder the users' competition or even result in task starvation. Although the second solution [23] adopts a multi-stage sampling-accepting process, at the first stage it uses Dynkin's algorithm [24] for the classic secretary problem. Therefore, this solution cannot guarantee the consumer sovereignty, as it is known that Dynkin's algorithm adopts a two-stage sampling-accepting process.

1) *Mechanism Design*: To address the above challenges, we design an online mechanism, *Homo-OMZ*, by using a *multiple-stage sampling-accepting* process. The mechanism dynamically increases the sample size and learns a *bid threshold* used for future decision, while increasing the *stage-task-number* it allocates at various stages. The whole process is illustrated in Algorithm 1. We first divide all of  $T$  time steps into  $(\lfloor \log_2 T \rfloor + 1)$  stages:  $\{1, 2, \dots, \lfloor \log_2 T \rfloor, \lfloor \log_2 T \rfloor + 1\}$ . The stage  $i$  ends at time step  $T' = \lfloor 2^{i-1} T / 2^{\lfloor \log_2 T \rfloor} \rfloor$ . Correspondingly, the stage-task-number for the  $i$ -th stage is set to be  $L' = 2^{i-1} L / 2^{\lfloor \log_2 T \rfloor}$ , meaning that  $L'$  tasks should be allocated before the end time of this stage. Finally,  $L$  tasks in total should be allocated before the deadline  $T$ . Fig. 2 is an illustration when  $T = 8$ . When a stage is over, we add all users who have arrived into the sample set  $S'$ , and compute a bid threshold  $b^*$  according to the information of samples and the allocated stage-task-number  $L'$ . This bid threshold is computed by calling the **GetBidThreshold1** algorithm

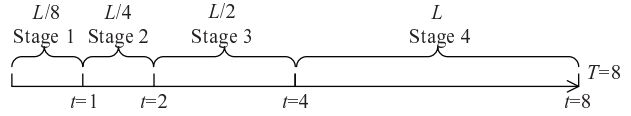


Fig. 2. Illustration of a multiple-stage sampling-accepting process when  $T = 8$ .

(to be elaborated later), and used for making decision at the next stage. Specially, when the last stage  $i = \lfloor \log_2 T \rfloor + 1$  comes, the bid threshold has been computed according to the information of all users arriving before time step  $\lfloor T/2 \rfloor$ , and the allocated stage-task-number  $L/2$ .

When a new user  $i$  arrives, the mechanism allocates a task to it as long as its bid is not less than the current bid threshold  $b^*$ , and the allocated stage-task-number  $L'$  has not been achieved. Meanwhile, we give user  $i$  a payment:

$$p_i = b^*,$$

and add this user to the set of selected users  $S$ . To start the mechanism, we initially set a large bid threshold  $\beta$ , which is used for making decision at the first stage.

Since each stage maintains a common bid threshold, it is natural to compute the lowest single price that many users will accept from the sample set  $S'$  and the allocated stage-task-number  $L'$  as the bid threshold. Similar to the solution of [23], we set the bid threshold equal to the  $L'$ -th lowest bid in the sample set, as illustrated in Algorithm 2.

2) *Mechanism Analysis*: In the following we analyze the properties of the *Homo-OMZ* mechanism from five aspects.

---

#### Algorithm 1: Homogeneous Online Mechanism under Zero Arrival-departure Interval Model (Homo-OMZ)

---

**Input:** Task number  $L$ , deadline  $T$   
1  $(t, T', L', S', b^*, S) \leftarrow (1, \frac{T}{2^{\lfloor \log_2 T \rfloor}}, \frac{L}{2^{\lfloor \log_2 T \rfloor}}, \emptyset, \beta, \emptyset)$ ;  
2 **while**  $t \leq T$  **do**  
3     **if** there is a user  $i$  arriving at time step  $t$  **then**  
4         **if**  $b_i \leq b^*$  **and**  $|S| < L'$  **then**  
5              $f_i \leftarrow 1$ ;  $p_i \leftarrow b^*$ ;  
6              $S \leftarrow S \cup \{i\}$ ;  
7         **else**  
8              $f_i \leftarrow 0$ ;  $p_i \leftarrow 0$ ;  
9         **end**  
10          $S' \leftarrow S' \cup \{i\}$ ;  
11     **end**  
12     **if**  $t = \lfloor T' \rfloor$  **then**  
13          $b^* \leftarrow \text{GetBidThreshold1}(L', S')$ ;  
14          $T' \leftarrow 2T'$ ;  $L' \leftarrow 2L'$ ;  
15     **end**  
16      $t \leftarrow t + 1$ ;  
17 **end**

---



---

#### Algorithm 2: GetBidThreshold1

---

**Input:** Stage-task-number  $L'$ , sample set  $S'$   
1 Sort  $S'$  s.t.  $b_1 \leq b_2 \leq \dots \leq b_{|S'|}$ ;  
2 **return**  $b_{L'}$ ;  


---

**Lemma 1.** *The HOMO-OMZ mechanism is computationally efficient.*

*Proof:* Since the mechanism runs online, we only need to focus on the computation complexity at each time step  $t \in \{1, \dots, T\}$ . It is easy to know that the running time of computing the allocation and payment of each user  $i$  (lines 3-11) is  $O(1)$ . Thus, the computation complexity of the mechanism is dominated by the complexity of computing the bid threshold (Algorithm 2). We can use any computationally efficient sorting algorithm to implement the Algorithm 2. For example, the computation complexity of quick sorting algorithm is  $O(n \log n)$ . Thus, the Homo-OMZ mechanism is computationally efficient. ■

**Lemma 2.** *The Homo-OMZ mechanism is individually rational.*

*Proof:* From the lines 4-9 of Algorithm 1, we know that  $p_i \geq b_i$  if  $i \in \mathcal{S}$ , otherwise  $p_i = 0$ . Thus, we have  $u_i \geq 0$ . ■

Designing a cost-truthful mechanism relies on the rationale of *bid-independence*. Let  $b_{-i}$  denote the sequence of bids arriving before the  $i$ -th bid  $b_i$ , i.e.,  $b_{-i} = (b_1, \dots, b_{i-1})$ . We call such a sequence *prefixal*. Let  $p'$  be a function from prefixal sequences to prices (non-negative real numbers). We extend the definition of bid-independence [25] to the online scenario:

**Definition 1** (Bid-independent Online Auction). *An online auction is called bid-independent if the allocation and payment rules for each player  $i$  satisfy:*

- a) *The auction constructs a price schedule  $p'(b_{-i})$ ;*
- b) *If  $p'(b_{-i}) \geq b_i$ , player  $i$  wins at price  $p_i = p'(b_{-i})$ ;*
- c) *Otherwise, player  $i$  is rejected, and  $p_i = 0$ .*

**Proposition 1.** ([20], Proposition 2.1) *An online auction is cost-truthful if and only if it is bid-independent.*

**Lemma 3.** *The Homo-OMZ mechanism is cost-truthful.*

*Proof:* Consider a user  $i$  that arrives at some stage for which the bid threshold is  $b^*$ . If by the time the user arrives there are no remaining tasks unallocated, then the user's cost declaration will not affect the allocation of the mechanism and thus cannot improve its utility by submitting a false cost. Otherwise, assume there are remaining tasks to be allocated by the time when the user arrives. In case  $c_i \leq b^*$ , reporting any cost below  $b^*$  would not make a difference in the user's allocation and payment, and its utility would be  $b^* - c_i \geq 0$ . Declaring a cost above  $b^*$  would make the worker lose the auction, and its utility would be 0. In case  $c_i > b^*$ , declaring any cost above  $b^*$  would leave the user unallocated with utility 0. If the user declares a cost lower than  $b^*$  it will be allocated. In such a case, however, its utility will be negative. Hence the user's utility is always maximized by reporting its true cost:  $b_i = c_i$ . ■

**Lemma 4.** *The Homo-OMZ mechanism satisfies the consumer sovereignty.*

*Proof:* Each stage is an accepting process as well as a sampling process ready for the next stage. As a result, users are not automatically rejected during the sampling process,

and are allocated as long as their bids are not less than the current bid threshold, and there are still remaining tasks unallocated at the current stage. ■

According to the proof of Theorem 2.1 from [23], it is easy to know that the expected total cost of the  $L$  users selected by the Homo-OMZ mechanism is at most  $1 + O(\sqrt{1/L})$  times the sum of the  $L$  lowest costs of all users. Thus, we obtain the following lemma.

**Lemma 5.** *The Homo-OMZ mechanism is  $O(1)$ -competitive for social efficiency.*

The above five lemmas together prove the following theorem.

**Theorem 1.** *The Homo-OMZ mechanism satisfies the computational efficiency, individual rationality, truthfulness, consumer sovereignty, and constant competitiveness for social efficiency under the homogeneous user model and zero arrival-departure interval model.*

## B. Heterogeneous User Mechanism

Although the Homo-OMZ mechanism satisfies four desirable properties (computational efficiency, individual rationality, truthfulness, and consumer sovereignty), and achieves constant competitiveness for social efficiency, the failure of guaranteeing good frugality makes it less attractive. To achieve good frugality, one possible direction is to make use of the off-the-shelf results on the budgeted feasible mechanisms [16]. The budgeted feasible mechanism design problem is similar to our frugal mechanism design problem, with the difference that the total payment paid to the winners is a constraint instead of an objective function. To address this issue, it is intuitive that we can dynamically learn a budget that are enough for allocating users for completing a specific number of tasks, then use this budget to compute a bid threshold by using budget feasible mechanisms, and finally use this bid threshold for making further decisions. In the following, we present a novel online mechanism *Hetero-OMZ* that satisfies all five desirable properties including the frugality in the heterogeneous user model. Note that since the homogeneous user model is a special case of the heterogeneous user model, the *Hetero-OMZ* mechanism is a general solution for both two models.

1) *Mechanism Design:* As illustrated in Algorithm 3, the basic framework of the *Hetero-OMZ* mechanism is very similar to the Homo-OMZ mechanism, with two differences: 1) each user may be allocated multiple tasks upper bounded by the number of tasks that it has announced, as long as its bid is not less than the current bid threshold  $b^*$ , and the allocated stage-task-number  $L'$  has not been achieved; 2) it has a different approach to computing the bid threshold, as illustrated in Algorithm 4.

To compute the bid threshold, we first find the minimal cost for performing  $\delta L'$  tasks from the sample set  $\mathcal{S}'$ . This can be done by using a simple greedy algorithm which sorts users according to their bids, and preferentially allocates tasks to users with lower bids until that all of  $\delta L'$  tasks have

---

**Algorithm 3: Heterogeneous Online Mechanism under Zero Arrival-departure Interval Model (*Hetero-OMZ*)**


---

**Input:** Task number  $L$ , deadline  $T$   
1  $(t, T', L', S', b^*, \mathcal{S}) \leftarrow (1, \frac{T}{2^{\lceil \log_2 T \rceil}}, \frac{L}{2^{\lceil \log_2 T \rceil}}, \emptyset, \beta, \emptyset)$ ;  
2 **while**  $t \leq T$  **do**  
3   **if** there is a user  $i$  arriving at time step  $t$  **then**  
4     **if**  $b_i \leq b^*$  **and**  $\sum_{j \in \mathcal{S}} f_j < L'$  **then**  
5        $f_i \leftarrow \min\{\tau_i, L' - \sum_{j \in \mathcal{S}} f_j\}$ ;  $p_i \leftarrow b^*$ ;  
6        $\mathcal{S} \leftarrow \mathcal{S} \cup \{i\}$ ;  
7     **else**  
8        $f_i \leftarrow 0$ ;  $p_i \leftarrow 0$ ;  
9     **end**  
10     $S' \leftarrow S' \cup \{i\}$ ;  
11   **end**  
12   **if**  $t = \lfloor T' \rfloor$  **then**  
13      $b^* \leftarrow \text{GetBidThreshold2}(L', S')$ ;  
14      $T' \leftarrow 2T'$ ;  $L' \leftarrow 2L'$ ;  
15   **end**  
16    $t \leftarrow t + 1$ ;  
17 **end**

---



---

**Algorithm 4: GetBidThreshold2**


---

**Input:** Stage-task-number  $L'$ , sample set  $S'$   
1  $\mathcal{J} \leftarrow \emptyset$ ;  $i \leftarrow \arg \min_{j \in S'} b_j$ ;  
2 **while**  $\sum_{j \in \mathcal{J}} f_j < \delta L'$  **do**  
3    $f_i \leftarrow \min\{\tau_i, \delta L' - \sum_{j \in \mathcal{J}} f_j\}$ ;  
4    $\mathcal{J} \leftarrow \mathcal{J} \cup \{i\}$ ;  
5    $i \leftarrow \arg \min_{j \in S' \setminus \mathcal{J}} b_j$ ;  
6 **end**  
7  $B \leftarrow \sum_{j \in \mathcal{J}} f_j b_j$ ;  
8  $p \leftarrow \text{BudgetFeasibleMechanism}(B, S')$ ;  
9 **return**  $p$ ;

---

been allocated (lines 1-6). Here we set  $\delta > 1$  to obtain a slight overestimate of the required budget for allocating  $L'$  tasks. Intuitively, if the value of  $\delta$  is too small, the budget will be not enough that the required number of tasks can be allocated at the next stage. Conversely, if the value of  $\delta$  is too large, the budget will be wasted, which will result in a bad frugality. Later we will fix the value of  $\delta$  elaborately to enable the mechanism achieving a constant frugality ratio. Second, we use the total payment for performing  $\delta L'$  tasks as the budget  $B$  (line 7), and then compute a bid threshold from the sample set  $S'$  and the budget  $B$  by using the budget feasible mechanism (line 8).

Note that the budget feasible mechanism is an offline procedure proposed by [16], which has access all bids. As illustrated in Algorithm 5, it also adopts a greedy strategy. First, users are sorted according to their bids so that:  $b_1 \leq b_2 \leq \dots \leq b_{|S'|}$ . When given a price at  $p = b_i$  and the remaining budget  $B - p \sum_{j \in \mathcal{J}} f_j$ , the number of tasks that a user could be allocated at this price without exceeding the budget constraint is  $f_i = \min\{\tau_i, \lfloor \frac{B}{p} \rfloor - \sum_{j \in \mathcal{J}} f_j\}$ . Then we find the largest  $k$  such that  $b_k \leq \frac{B}{\sum_{j=1}^{k-1} f_j + 1}$ . The set of selected users is  $\mathcal{J} = \{1, 2, \dots, k\}$ . Finally, the bid threshold is set to be  $b_k$ . This mechanism follows the *proportional share allocation rule* in essence. It has a good property that it sets a common price which on one hand

is high enough so that enough users could accept and on the other hand is low enough so that the budget could be efficiently exploited.

---

**Algorithm 5: BudgetFeasibleMechanism [16]**


---

**Input:** Budget constraint  $B$ , sample set  $S'$   
1  $\mathcal{J} \leftarrow \emptyset$ ;  $i \leftarrow \arg \min_{j \in S'} b_j$ ;  
2 **while**  $b_i \leq \frac{B}{\sum_{j \in \mathcal{J}} f_j + 1}$  **do**  
3    $p \leftarrow b_i$ ;  
4    $f_i \leftarrow \min\{\tau_i, \lfloor \frac{B}{p} \rfloor - \sum_{j \in \mathcal{J}} f_j\}$ ;  
5    $\mathcal{J} \leftarrow \mathcal{J} \cup \{i\}$ ;  
6    $i \leftarrow \arg \min_{j \in S' \setminus \mathcal{J}} b_j$ ;  
7 **end**  
8 **return**  $p$ ;

---

In the following, we use an example to illustrate how the *Hetero-OMZ* mechanism works.

**Example 1.** Consider a crowdsourcer with the number of tasks to perform  $L = 8$  and the deadline  $T = 8$ . There are five users arriving online before the deadline with the following profiles:  $\theta_1 = (1, 1, 4, 2)$ ,  $\theta_2 = (2, 2, 4, 4)$ ,  $\theta_3 = (4, 4, 4, 5)$ ,  $\theta_4 = (6, 6, 4, 1)$ , and  $\theta_5 = (7, 7, 4, 3)$ .

We set  $\beta = 5$  and  $\delta = 2$ . Then the *Hetero-OMZ* mechanism works as follows.

- ◇  $t = 1$ :  $(T', L', S', b^*, \mathcal{S}) = (1, 1, \emptyset, 5, \emptyset)$ ,  $f_1 = 1$ ,  $p_1 = 5$ ,  $\mathcal{S} = \{1\}$ ,  $S' = \{1\}$ . Update the bid threshold:  $b^* = 2$ .
- ◇  $t = 2$ :  $(T', L', S', b^*, \mathcal{S}) = (2, 2, \{1\}, 2, \{1\})$ ,  $f_2 = 0$ ,  $p_2 = 0$ ,  $\mathcal{S} = \{1\}$ ,  $S' = \{1, 2\}$ . Update the bid threshold:  $b^* = 2$ .
- ◇  $t = 4$ :  $(T', L', S', b^*, \mathcal{S}) = (4, 4, \{1, 2\}, 2, \{1\})$ ,  $f_3 = 0$ ,  $p_3 = 0$ ,  $\mathcal{S} = \{1\}$ ,  $S' = \{1, 2, 3\}$ . Update the bid threshold:  $b^* = 4$ .
- ◇  $t = 6$ :  $(T', L', S', b^*, \mathcal{S}) = (8, 8, \{1, 2, 3\}, 4, \{1\})$ ,  $f_4 = 4$ ,  $p_4 = 4$ ,  $\mathcal{S} = \{1, 4\}$ ,  $S' = \{1, 2, 3, 4\}$ .
- ◇  $t = 7$ :  $(T', L', S', b^*, \mathcal{S}) = (8, 8, \{1, 2, 3, 4\}, 4, \{1, 4\})$ ,  $f_5 = 3$ ,  $p_5 = 4$ ,  $\mathcal{S} = \{1, 4, 5\}$ ,  $S' = \{1, 2, 3, 4, 5\}$ . Finally, the set of selected users is  $\mathcal{S} = \{1, 4, 5\}$ , and the payments to these selected 3 users are 5, 16, and 12 respectively.

2) *Mechanism Analysis*: It can be proved that the *Hetero-OMZ* mechanism also satisfies the computational efficiency, individual rationality, truthfulness, and consumer sovereignty, with very similar proof of the *Homo-OMZ* mechanism. In the following we prove that the *Hetero-OMZ* mechanism can achieve a constant frugality ratio under both the i.i.d. model and the secretary model by elaborately fixing different value of  $\delta$  in Algorithm 4.

If the stage-task-number could be achieved at each stage, then  $L$  tasks would be allocated finally. Since our *Hetero-OMZ* mechanism consists of multiple stages, and dynamically increases the stage-task-number, it only needs to prove that  $L/2$  tasks could be allocated at the last stage while the total payment is no more than the budget  $B$ . In this case, the frugality ratio would be  $\delta$ , since at the last stage the budget  $B$  is the minimal cost for



performing  $\delta L' = \delta L/2$  tasks according to Algorithm 4. In order to facilitate analysis, we change the stage-task-number constraint into the budget constraint at each stage in Algorithm 3. Specially, in line 4 of Algorithm 3, the condition  $\sum_{j \in \mathcal{S}} f_j < L'$  is replaced with  $\sum_{j \in \mathcal{S}} p_j f_j < B$ , and in line 5, the allocation  $f_i \leftarrow \min\{\tau_i, L' - \sum_{j \in \mathcal{S}} f_j\}$  is replaced with  $f_i \leftarrow \min\{\tau_i, \lfloor (B - \sum_{j \in \mathcal{S}} p_j f_j) / b^* \rfloor\}$ . Note that in this case, if we prove that at least  $L/2$  tasks could be allocated at the last stage under the budget constraint  $B$ , then it is equivalent to that  $L/2$  tasks could be allocated while the total payment is no more than  $B$ , meaning that the frugality ratio is  $\delta$ . In the following analysis, we use the modified version of Algorithm 3.

For ease of illustration, we first introduce more notations and concepts.

- $\mathcal{Z}$  denotes the set of selected users computed by Algorithm 5 based on the set of users  $\mathcal{U}$  arriving before  $T$  and the budget  $2B$ . The bid threshold of  $\mathcal{Z}$  is  $p$ .
- $\mathcal{S}'$  denotes the sample set obtained when the stage  $\lfloor \log_2 T \rfloor$  is over, which consists of all users arriving before the time  $\lfloor T/2 \rfloor$ .
- $\mathcal{Z}_1$  and  $\mathcal{Z}_2$  denote the subsets of  $\mathcal{Z}$  that appears in the first and second half of the input stream, respectively. Thus we have  $\mathcal{Z}_1 = \mathcal{Z} \cap \mathcal{S}'$ , and  $\mathcal{Z}_2 = \mathcal{Z} \cap \{\mathcal{U} \setminus \mathcal{S}'\}$ .
- $\mathcal{Z}'_1$  denotes the set of selected users computed by Algorithm 5 based on the sample set  $\mathcal{S}'$  and the budget  $B$ . The bid threshold of  $\mathcal{Z}'_1$  is  $p'_1$ .
- $\mathcal{Z}'_2$  denotes the set of selected users computed by Algorithm 3 at the last stage.
- The total number of tasks allocated to a set of selected users  $\mathcal{X}$  is denoted by a function  $f(\mathcal{X}) = \sum_{i \in \mathcal{X}} f_i$ .

Before analyzing the frugality formally, we introduce a lemma about Algorithm 5.

**Lemma 6.** ([16], Lemma 3.1) *For a given sample set of users, let  $L$  be the maximal number of tasks that can be allocated under a given budget. Then at least  $L/2$  tasks can be allocated under budget at the price computed by Algorithm 5.*

According to Lemma 6, we have  $f(\mathcal{Z}'_1) \geq \delta L'/2 = \delta L/4$ .

**Lemma 7.** *When  $\delta = 2$ , we have  $\mathbb{E}[f(\mathcal{Z}'_2)] \geq L/2$  under the i.i.d. model.*

The proof of Lemma 7 is given in Appendix A.

Different from the i.i.d. model, under the secretary model we make  $\mathcal{Z}$  denote the set of selected users computed by Algorithm 5 based on the set of users  $\mathcal{U}$  arriving before  $T$  and the budget  $B$ . Other notations and concepts remain unchanged. In addition, it is assumed that the number of tasks that each user can complete is at most  $f(\mathcal{Z})/\omega$ . In order to facilitate analysis, we introduce a lemma as follows.

**Lemma 8.** ([26], Lemma 16) *For sufficiently large  $\omega$ , the random variable  $|f(\mathcal{Z}_1) - f(\mathcal{Z}_2)|$  is bounded by  $f(\mathcal{Z})/2$  with a constant probability.*

Because  $f(\mathcal{X})$  is a linear function, we have  $f(\mathcal{Z}_1) +$

$f(\mathcal{Z}_2) = f(\mathcal{Z})$ . Thus, Lemma 8 can be easily extended to the following corollary.

**Corollary 1.** *For sufficiently large  $\omega$ , both  $f(\mathcal{Z}_1)$  and  $f(\mathcal{Z}_2)$  are at least  $f(\mathcal{Z})/4$  with a constant probability.*

**Lemma 9.** *For sufficiently large  $\omega$ , when  $\delta = 8$ , we have  $f(\mathcal{Z}'_2) \geq L/2$  with a constant probability under the secretary model.*

The proof of Lemma 9 is given in Appendix B. The above analysis proves the following theorem.

**Theorem 2.** *The Hetero-OMZ mechanism satisfies the computational efficiency, individual rationality, truthfulness, consumer sovereignty, and constant frugality under the zero arrival-departure interval model.*

#### IV. ONLINE MECHANISM UNDER GENERAL INTERVAL MODEL

In this section, we consider the general interval model, and there may be multiple online users in the auction simultaneously. Firstly, we change the settings of Example 1 to show that the *Hetero-OMZ* mechanism is not time-truthful in the general interval model.

**Example 2.** *All the settings are the same as Example 1 except for that user 1 has a non-zero arrival-departure interval,  $a_1 < d_1$ . Specially, the profile of user 1 is  $\theta_1 = (1, 5, 4, 2)$ .*

In this example, if user 1 report its profile truthfully, then it will obtain the payment 5 according to the *Hetero-OMZ* mechanism. However, if user 1 delays announcing its arrival time and reports  $\theta'_1 = (5, 5, 4, 2)$ , then it will improve its payment to 20 according to the *Hetero-OMZ* mechanism (the detailed computing process is omitted).

In the following, we will present a new online mechanism, *Hetero-OMG*, and prove that it satisfies all five desirable properties in the general interval model.

##### A. Mechanism Design

Since the *Hetero-OMZ* mechanism can be applied in both the homogeneous user model and the heterogeneous user model, and can satisfy several desirable properties, we adopt a similar algorithm framework under the general interval model. Meanwhile, in order to guarantee the *cost*- and *time-truthfulness*, it is necessary to modify the *Hetero-OMZ* mechanism based on three principles. First, any user is added to the sample set only when it departs; otherwise, the bid-independence will be destroyed if its arrival-departure time spans multiple stages, because a user can indirectly affect its payment now. Second, if there are multiple users who have not yet departed at some time, we sort these online users according to the numbers of tasks that they can complete, instead of their bids, and preferentially select those users with higher number of tasks. In this way, the bid-independence can be held. Third, whenever a new time step arrives, it scans through the list of users who have not yet departed and allocates tasks



to those whose bids are not larger than the current bid threshold under the stage-task-number constraint, even if some arrived much earlier. At the departure time of any user who was selected as a winner, the user is paid for a payment equal to the maximum payment attained during the user's reported arrival-departure interval, even if this payment is larger than the payment at the time step when the user was selected as a winner.

---

**Algorithm 6:** Heterogeneous Online Mechanism under General Interval Model (*Hetero-OMG*)

---

**Input:** Task Number  $L$ , deadline  $T$

```

1  $(t, T', L', S', b^*, \mathcal{S}) \leftarrow (1, \frac{T}{2^{\lceil \log_2 T \rceil}}, \frac{L}{2^{\lceil \log_2 T \rceil}}, \emptyset, \beta, \emptyset);$ 
2 while  $t \leq T$  do
3   Add all new users arriving at time step  $t$  to a set of
   online users  $\mathcal{O}$ ;  $\mathcal{O}' \leftarrow \mathcal{O} \setminus S$ ;
4   repeat
5      $i \leftarrow \arg \max_{j \in \mathcal{O}'} \tau_j$ ;
6     if  $b_i \leq b^*$  and  $\sum_{j \in S} f_j < L'$  then
7        $f_i \leftarrow \min\{\tau_i, L' - \sum_{j \in S} f_j\}$ ;  $p_i \leftarrow b^*$ ;
8        $S \leftarrow S \cup \{i\}$ ;
9     else
10       $f_i \leftarrow 0$ ;  $p_i \leftarrow 0$ ;
11    end
12     $\mathcal{O}' \leftarrow \mathcal{O}' \setminus \{i\}$ ;
13  until  $\mathcal{O}' = \emptyset$ ;
14  Remove all users departing at time step  $t$  from  $\mathcal{O}$ , and
  add them to  $S'$ ;
15  if  $t = \lfloor T' \rfloor$  then
16     $b^* \leftarrow \text{GetBidThreshold2}(L', S')$ ;
17     $T' \leftarrow 2T'$ ;  $L' \leftarrow 2L'$ ;  $\mathcal{O}' \leftarrow \mathcal{O}$ ;
18    repeat
19       $i \leftarrow \arg \max_{j \in \mathcal{O}'} \tau_j$ ;
20      if  $b_i \leq b^*$  and
       $\min\{\tau_i, L' + f_i - \sum_{j \in S} f_j\} b^* > f_i p_i$  then
21         $f_i \leftarrow \min\{\tau_i, L' + f_i - \sum_{j \in S} f_j\}$ ;
22         $p_i \leftarrow b^*$ ;
23        if  $i \notin S$  then  $S \leftarrow S \cup \{i\}$ ;
24      end
25       $\mathcal{O}' \leftarrow \mathcal{O}' \setminus \{i\}$ ;
26    until  $\mathcal{O}' = \emptyset$ ;
27   $t \leftarrow t + 1$ ;
28 end
```

---

According to the above principles, we design the *Hetero-OMG* mechanism satisfying all desirable properties under the general interval model, as described in Algorithm 6. Specially, we consider two cases. The first case is when the current time step  $t$  is not at the end of any stage. In this case, the bid threshold remains unchanged. The following operations (the lines 3-14 in Algorithm 6) are performed. First, all new users arriving at time step  $t$  are added to a set of online users  $\mathcal{O}$ . Then we make decision on whether to select these online users one by one in the order of the numbers of tasks that they can complete; the users who can complete more tasks will be selected first. If an online user  $i$  has been selected as a winner before time step  $t$ , we need not to make decision on it again because it is impossible to obtain a higher payment than before (to be proved later in Appendix C). Otherwise, we need to make decision on it again: it will be allocated multiple tasks upper bounded

by the number of tasks that it has announced, as long as its bid is not less than the current bid threshold  $b^*$ , and the allocated stage-task-number  $L'$  has not been achieved; meanwhile, we pay user  $i$  at price  $p_i = b^*$  per task, and add it to the set of selected users  $S$ . Finally, we remove all users departing at time step  $t$  from  $\mathcal{O}$ , and add them to the sample set  $S'$ .

The second case is when the current time step is just at the end of some stage. In this case, the bid threshold will be updated. The mechanism works as the lines 16-25. We need to make decision on whether to select these online users, and at what prices, one by one in the order of the numbers of tasks that they can complete, no matter whether they have ever been selected as the winners before time step  $t$ . As shown in the lines 20-23, if user  $i$  can obtain a higher payment than before (this could be because that this user is given a higher price or is allocated more tasks), its price or number of tasks allocated will be updated. Meanwhile, if user  $i$  has never been selected as a winner before time step  $t$ , it will be added to the set  $S$ .

Return to Example 2. If all of the five users report their types truthfully, then the *Hetero-OMG* mechanism works as follows.

- ◊  $t = 1$ :  $(T', B', S', b^*, \mathcal{S}) = (1, 1, \emptyset, 5, \emptyset)$ ,  $f_1 = 1$ ,  $p_1 = 5$ ,  $S = \{1\}$ ,  $S' = \emptyset$ . Update the bid threshold:  $b^* = 5$ , update the allocation and payment of user 1:  $f_1 = 2$ ,  $p_1 = 5$ .
- ◊  $t = 2$ :  $(T', B', S', b^*, \mathcal{S}) = (2, 2, \emptyset, 5, \{1\})$ ,  $f_2 = 0$ ,  $p_2 = 0$ ,  $S = \{1\}$ ,  $S' = \{2\}$ . Update the bid threshold:  $b^* = 4$ , update the allocation and payment of user 1:  $f_1 = 4$ ,  $p_1 = 4$ .
- ◊  $t = 4$ :  $(T', B', S', b^*, \mathcal{S}) = (4, 4, \{2\}, 4, \{1\})$ ,  $f_3 = 0$ ,  $p_3 = 0$ ,  $S = \{1\}$ ,  $S' = \{2, 3\}$ . Update the bid threshold:  $b^* = 5$ , update the allocation and payment of user 1:  $f_1 = 4$ ,  $p_1 = 5$ .
- ◊  $t = 5$ : user 1 departs, so  $S' = \{1, 2, 3\}$ .
- ◊  $t = 6$ :  $(T', B', S', b^*, \mathcal{S}) = (8, 8, \{1, 2, 3\}, 5, \{1\})$ ,  $f_4 = 4$ ,  $p_4 = 5$ ,  $S = \{1, 4\}$ ,  $S' = \{1, 2, 3, 4\}$ . Now all of 8 tasks have been allocated.
- ◊  $t = 7$ :  $(T', B', S', b^*, \mathcal{S}) = (8, 8, \{1, 2, 3, 4\}, 5, \{1, 4\})$ ,  $f_5 = 0$ ,  $p_5 = 0$ ,  $S = \{1, 4\}$ ,  $S' = \{1, 2, 3, 4, 5\}$ .

Thus, user 1 can obtain the payment 20 according to the *Hetero-OMG* mechanism. Even if user 1 delays announcing its arrival time and reports  $\theta'_1 = (5, 5, 4, 2)$ , it still cannot improve its payment (the detailed computing process is omitted). Therefore, the time-truthfulness can be guaranteed in the general interval model.

### B. Mechanism Analysis

It is convenient to prove that the *Hetero-OMG* mechanism also satisfies the *computational efficiency*, *individual rationality*, *consumer sovereignty*, and *constant frugality* as *Hetero-OMZ* (with almost the same proof), although *Hetero-OMG* may have slightly higher frugality ratio than *Hetero-OMZ*. Most importantly, we can prove the truthfulness of the *Hetero-OMG* mechanism.

**Lemma 10.** *The Hetero-OMG mechanism is cost- and time-truthful.*

The proof of Lemma 10 is given in Appendix C.

**Theorem 3.** *The Hetero-OMG mechanism satisfies the computational efficiency, individual rationality, budget feasibility, truthfulness, consumer sovereignty, and constant frugality under the general interval model.*

## V. PERFORMANCE EVALUATION

To evaluate the performance of our online mechanisms, we implemented *Homo-OMZ*, *Hetero-OMZ* and *Hetero-OMG*, and compared them against the following two benchmarks. The first benchmark is the *optimal* offline solution which has full knowledge about all users' profiles. This can be done by using a simple greedy algorithm as illustrated in Algorithm 4 (lines 1-6), which sorts users according to their bids, and preferentially allocates tasks to users with lower bids until that all of tasks have been allocated. The second benchmark is the *random* mechanism, which adopts a naive strategy, i.e., rewards users based on an uninformed fixed bid threshold.

### A. Evaluation under Homogeneous User Model and Zero Arrival-departure Interval Model

We first evaluate the *Homo-OMZ* and *Hetero-OMZ* mechanisms under the homogeneous user model and zero arrival-departure interval model.

**Simulation Setup:** We set the deadline to  $T=1800s$ , and vary the number  $L$  of tasks to be completed from 100 to 400 with the increment of 100. Users arrive according to a Poisson process in time with arrival rate  $\lambda = 0.6$ . For each user  $i \in \mathcal{U}$ , we set that:  $a_i = d_i$ ,  $\tau_i = 1$ ,  $c_i \sim U[1, 10]$ , where  $U[a, b]$  denotes a uniform distribution over  $[a, b]$ . For the *Hetero-OMZ* mechanism, we set  $\delta = 1$  and  $\delta = 2$  respectively, and  $\beta = 10$ . As we proved in Lemma 7, when  $\delta = 2$  the *Hetero-OMZ* mechanism is constant-frugal. Here we also set  $\delta = 1$  for comparison. For the *optimal* offline mechanism, we compute the minimum cost of performing  $L$  and  $2L$  tasks in order to obtain frugality ratios under the two frameworks of “idealistic frugality” and “realistic frugality”, respectively. For the *random* mechanism, we obtain the average performance of 50 such solutions for evaluations, where in each solution the bid threshold is chosen at random from the range of 1 to 10.

**Simulation Results:** Fig. 3(a) compares the crowd-sourcer's total payment required by the *Homo-OMZ* and *Hetero-OMZ* mechanisms against the two benchmarks. Fig. 3(b) compares the number of tasks completed by the three online mechanisms. From the simulation results, we can observe the following four phenomena:

- The total payments of all evaluated mechanisms increase with the number of tasks, and the growth rate is larger than 1 (except for the *random* mechanism). It is because that the user set is limited, and the users with higher costs must be selected if more tasks are required to be completed.

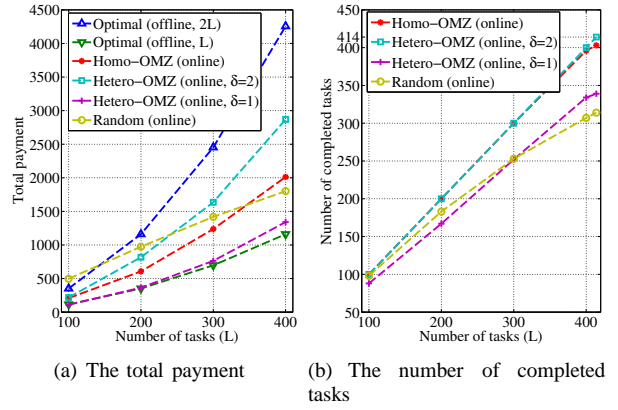


Fig. 3. Evaluation results under the homogeneous user model and zero arrival-departure interval model.

- Both of the *Homo-OMZ* and *Hetero-OMZ* mechanisms outweigh the *random* mechanism. Note that, although the *random* mechanism has less total payment under some cases, it has many tasks uncompleted, and the average price per task is higher than that of the other mechanisms.
- Only the *Hetero-OMZ* mechanism with  $\delta = 2$  can complete all of required tasks under various cases, although it requires higher total payment than the *Homo-OMZ* mechanism and the *Hetero-OMZ* mechanism with  $\delta = 1$ . Note that, at most 828 tasks can be completed (by the optimal offline mechanism) in our simulations, and a half of these tasks (414 tasks) can be completed by the *Hetero-OMZ* mechanism with  $\delta = 2$  (see Fig. 3(b)). In fact, it implies a trade-off between the frugality and the number of completed tasks, which should be considered by the mechanism designer. Anyway, the *Hetero-OMZ* mechanism is a good choice to guarantee the completion of required tasks without sacrificing the frugality.
- The total payment of the *Hetero-OMZ* mechanism with  $\delta = 2$  is less than the *optimal* offline mechanism with  $2L$  tasks, indicating that the “realistic” frugality ratio is less than 2, which is consistent with our theoretical analysis in Lemma 7. Moreover, by comparing the total payments of the *Hetero-OMZ* mechanism with  $\delta = 2$  and the *optimal* offline mechanism with  $L$  tasks, we have that the “idealistic” frugality ratio is less than 2.5. Anyway, it has been verified that the *Hetero-OMZ* mechanism can guarantee the constant frugality.

### B. Evaluation under Heterogeneous User Model

Now we evaluate the *Hetero-OMZ* and *Hetero-OMG* mechanisms under the heterogeneous user model.

**Simulation Setup:** We set the deadline to  $T=1800s$ , and vary  $L$  from 100 to 1000 with the increment of 100. Users arrive according to a Poisson process in time with arrival rate  $\lambda$ , and we vary  $\lambda$  from 0.2 to 1 with the increment of 0.2. For each user  $i \in \mathcal{U}$ , we set that:  $\tau_i \sim U[1, 10]$ ,  $c_i \sim U[1, 10]$ . For the *Hetero-OMZ* mechanism, each user

has zero arrival-departure interval. For the *Hetero-OMG* mechanism, the arrival-departure interval of each user is uniformly distributed over  $[0, 300]$  seconds. For both of the above two mechanisms, we set  $\delta = 2$  and  $\beta = 10$ . The settings of the *optimal* offline mechanism and the *random* mechanism are the same as in Section V-A.

**Simulation Results:** Fig. 4 compares the crowdsourcer's total payment required by the *Hetero-OMZ* and *Hetero-OMG* mechanisms against the two benchmarks. Fig. 5 compares the average price per task required by various mechanisms. From the simulation results, we can observe the following four phenomena:

- From Fig. 4(a) and Fig. 5(a), we can observe that all evaluated mechanisms (except for the *random* mechanism) require less total payment and average price per task when more users participate. From Fig. 4(b) and Fig. 5(b), we can observe that all evaluated mechanisms require higher total payment and average price per tasks (except for the *random* mechanism) when the number of tasks increases.
- Both of the *Hetero-OMZ* and *Hetero-OMG* mechanisms outweigh the *random* mechanism in terms of both the total payment and average price per task. Note that, although the *random* mechanism has less total payment or average price per task under some cases, it has many tasks uncompleted. By contrast, both of the other mechanisms can complete all required tasks.
- The total payment of the *Hetero-OMZ* mechanism is less than the *optimal* offline mechanism with  $2L$  tasks, indicating that the “realistic” frugality ratio is less than 2, which is consistent with our theoretical analysis in Lemma 7. Moreover, by comparing the total payments of the *Hetero-OMZ* mechanism and the *optimal* offline mechanism with  $L$  tasks, we have that the “idealistic” frugality ratio is less than 2.4. Anyway, it has been verified that the *Hetero-OMZ* mechanism can guarantee the constant frugality.
- The total payment of the *Hetero-OMG* mechanism is very close to the *optimal* offline mechanism with  $2L$  tasks, indicating that the “realistic” frugality ratio is close to 2. Besides, the *Hetero-OMG* mechanism requires higher total payment and average price per tasks than the *Hetero-OMZ* mechanism in order to guarantee the time-truthfulness.

## VI. RELATED WORK

### A. Mechanism Design for Mobile Crowd Sensing

Reddy et al. [27] developed recruitment frameworks to enable the crowdsourcer to identify well-suited participants for data collections. However, they focused only on the user selection instead of the incentive mechanism design. At present, there are some studies [11]–[15] on incentive mechanism design for MCS applications in the offline scenario. Generally, two system models are considered: the platform/crowdsourcer-centric model where the crowdsourcer provides a fixed reward to participating users, and the user-centric model where users can have their reserve

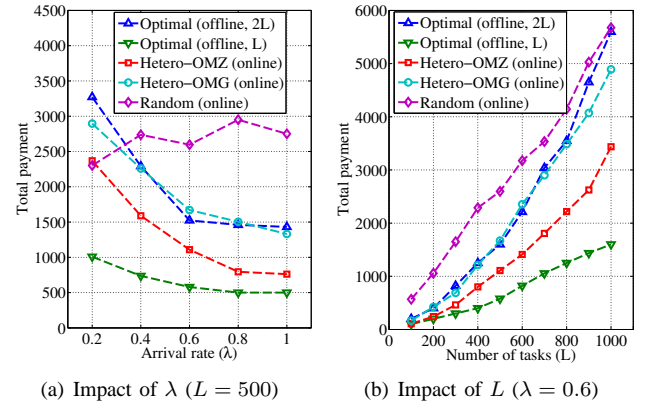


Fig. 4. The total payment under the heterogeneous user model.

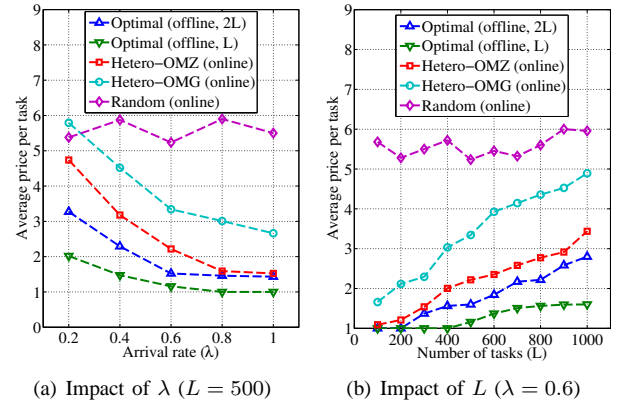


Fig. 5. The average price per task under the heterogeneous user model.

prices for the sensing service. For the crowdsourcer-centric model, incentive mechanisms were designed by using a Stackelberg game [13], [14]. The Nash Equilibrium and Stackelberg Equilibrium were computed as the solution, where the costs of all users or their probability distribution was assumed to be known. In contrast, the user-centric model allows that each user has a private cost only known to itself. Danezis et al. [11] developed a sealed-bid second-price auction to estimate the users' value of sensing data with location privacy. Lee and Hoh [12] designed and evaluated a reverse auction based dynamic price incentive mechanism, where users can sell their sensed data to a service provider with users' claimed bids. Jaimes et al. [15] proposed a recurrent reverse auction incentive mechanism with a greedy algorithm that selects a representative subset of the users according to their location given a fixed budget. Yang et al. [14] designed an auction-based incentive mechanism, and proved this mechanism was computationally efficient, individually rational, profitable, and truthful. However, all of these studies failed to account for the online arrival of users.

Recently, some researchers have begun to focus on online mechanism design for crowdsourcing markets [16]–[19], [28]. Singer et al. [16] and Singla et al. [17] presented incentive mechanisms for maximizing a linear utility function based on the bidding model and the posted price

model respectively. Badanidiyuru et al. [18] and Zhao et al. [19] considered incentive mechanisms for maximizing a submodular utility function. However, all of the above studies mainly focus on budget feasible mechanisms. Zhang et al. [28] considered incentive mechanisms for maximizing the platform utility, which is defined as the total value of selected users minus the total cost. Only Singer et al. [16] considered the frugal mechanisms partially, but it failed to consider the *consumer sovereignty* and *time-truthfulness*.

### B. Online Auctions and Generalized Secretary Problems

*Online auction* is the essence of many networked markets, in which information about goods, agents, and outcomes is revealed one by one online in a random order, and the agents must make irrevocable decisions without knowing future information. Combining optimal stopping theory with game theory provides us a powerful tool to model the actions of rational agents in an online auction. The theory of *optimal stopping* is concerned with the problem of choosing a time to take a particular action, in order to maximize an expected reward or minimize an expected cost. A classic problem of optimal stopping theory is the *secretary problem*: designing an algorithm for hiring one secretary from a pool of  $n$  applicants arriving online, to maximize the probability of hiring the best secretary [24]. Many variants of the classic secretary problem have been studied in the literature and the most relevant to this work is the *k-choice secretary problem*, in which the interviewer is allowed to hire up to  $k \geq 1$  applicants in order to maximize performance of the secretarial group based on their overlapping skills. Kleinberg [23] and Babaioff et al. [22] presented two constant competitive algorithms for a special *k-choice secretary problem* in which the objective function is a linear one, equaling to the sum of the individual values of selected applicants. They could also be used for minimizing the total cost of selected applicants in essence. However, they failed to consider the *truthfulness* and *consumer sovereignty*. Although some solutions ([21], [23], [29]) of online auctions provided good ideas of designing truthful mechanisms, they still could not satisfy the *consumer sovereignty*. Moreover, all of these solutions only applied to the homogeneous user model instead of the heterogeneous user model, and could only satisfy the social efficiency instead of the frugality.

## VII. CONCLUSIONS AND FUTURE WORK

In this paper, we have investigated online incentive mechanisms for mobile crowd sensing. We focus on frugal mechanisms which aim at minimizing the total payment while a specific number of tasks can be completed. We have designed three online mechanisms that are applicable to different models and hold different properties. The *Homo-OMZ* mechanism is applicable to the homogeneous user model and can satisfy the *social efficiency* but not *constant frugality*. The *Hetero-OMZ* and *Hetero-OMG* mechanisms are applicable to both the homogeneous and heterogeneous user models, and can satisfy the *constant frugality*.

The *Hetero-OMG* mechanism can also satisfy the *time-truthfulness* under a general interval model. Besides, all of these three mechanisms can satisfy the *computational efficiency*, *individual rationality*, *cost-truthfulness*, and *consumer sovereignty*.

An interesting open problem is to design frugal online incentive mechanisms for more complex scenarios. For example, each user can complete a subset of tasks, and the crowdsourcer wants to complete the whole set of tasks, or obtain a specific value from selected users where the value function is submodular.

## REFERENCES

- [1] G. Chatzimilioudis, A. Konstantinidis, C. Laoudias, and D. Zeinalipour-Yazti, "Crowdsourcing with smartphones," *IEEE Internet Comp.*, pp. 36–44, 2012.
- [2] "Sensorly," <http://www.sensorly.com>.
- [3] E. Koukoumidis, L.-S. Peh, and M. R. Martonosi, "Signalguru: leveraging mobile phones for collaborative traffic signal schedule advisory," in *Proc. ACM MobiSys*, 2011, pp. 127–140.
- [4] P. Mohan, V. N. Padmanabhan, and R. Ramjee, "Nericell: rich monitoring of road and traffic conditions using mobile smartphones," in *Proc. ACM SenSys*, 2008, pp. 323–336.
- [5] A. Thiagarajan, L. Ravindranath, K. LaCurts, S. Madden, H. Balakrishnan, S. Toledo, and J. Eriksson, "Vtrack: accurate, energy-aware road traffic delay estimation using mobile phones," in *Proc. ACM SenSys*, 2009, pp. 85–98.
- [6] R. K. Rana, C. T. Chou, S. S. Kanhere, N. Bulusu, and W. Hu, "Earphone: an end-to-end participatory urban noise mapping system," in *Proc. ACM/IEEE IPSN*, 2010, pp. 105–116.
- [7] M. Stevens and E. DHondt, "Crowdsourcing of pollution data using smartphones," in *Proc. ACM UbiComp*, 2010, pp. 1–4.
- [8] Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: wireless indoor localization with little human intervention," in *Proc. ACM MobiCom*, 2012, pp. 269–280.
- [9] N. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. Campbell, "A survey of mobile phone sensing," *IEEE Commun. Mag.*, vol. 48, no. 9, pp. 140–150, 2010.
- [10] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: Current state and future challenges," *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 32–39, 2011.
- [11] G. Danezis, S. Lewis, and R. Anderson, "How much is location privacy worth," in *Proc. WEIS*, 2005.
- [12] J. Lee and B. Hoh, "Sell your experiences: a market mechanism based incentive for participatory sensing," in *Proc. IEEE PerCom*, 2010, pp. 60–68.
- [13] L. Duan, T. Kubo, K. Sugiyama, J. Huang, T. Hasegawa, and J. Walrand, "Incentive mechanisms for smartphone collaboration in data acquisition and distributed computing," in *Proc. IEEE INFOCOM*, 2012, pp. 1701–1709.
- [14] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing," in *Proc. ACM MobiCom*, 2012, pp. 173–184.
- [15] L. Jaimes, I. Vergara-Laurens, and M. Labrador, "A location-based incentive mechanism for participatory sensing systems with budget constraints," in *Proc. IEEE PerCom*, 2012, pp. 103–108.
- [16] Y. Singer and M. Mittal, "Pricing mechanisms for crowdsourcing markets," in *Proc. WWW*, 2013, pp. 1157–1166.
- [17] A. Singla and A. Krause, "Truthful incentives in crowdsourcing tasks using regret minimization mechanisms," in *Proc. WWW*, 2013, pp. 1167–1178.
- [18] A. Badanidiyuru, R. Kleinberg, and Y. Singer, "Learning on a budget: posted price mechanisms for online procurement," in *Proc. ACM EC*, 2012, pp. 128–145.
- [19] D. Zhao, X.-Y. Li, and H. Ma, "How to crowdsource tasks truthfully without sacrificing utility: online incentive mechanisms with budget constraint," in *Proc. of IEEE INFOCOM*, 2014.
- [20] Z. Bar-Yossef, K. Hildrum, and F. Wu, "Incentive-compatible online auctions for digital goods," in *Proc. ACM-SIAM SODA*, 2002, pp. 964–970.
- [21] M. T. Hajiaghayi, R. Kleinberg, and D. C. Parkes, "Adaptive limited-supply online auctions," in *Proc. ACM EC*, 2004, pp. 71–80.

- [22] M. Babaioff, N. Immorlica, D. Kempe, and R. Kleinberg, "A knapsack secretary problem with applications," in *APPROX-RANDOM*, 2007, pp. 16–28.
- [23] R. Kleinberg, "A multiple-choice secretary algorithm with applications to online auctions," in *Proc. ACM-SIAM SODA*, 2005, pp. 630–631.
- [24] E. B. Dynkin, "The optimum choice of the instant for stopping a markov process," *Sov. Math. Dokl.*, vol. 4, pp. 627–629, 1963.
- [25] A. V. Goldberg, J. D. Hartline, A. R. Karlin, M. Saks, and A. Wright, "Competitive auctions," *Games and Economic Behavior*, vol. 55, no. 2, pp. 242–269, 2006.
- [26] M. Bateni, M. Hajiaghayi, and M. Zadimoghaddam, "Submodular secretary problem and extensions," in *APPROX-RANDOM*, 2010, pp. 39–52.
- [27] S. Reddy, D. Estrin, and M. Srivastava, "Recruitment framework for participatory sensing data collections," in *Proc. Pervasive*, 2010, pp. 138–155.
- [28] X. Zhang, Z. Yang, Z. Zhou, H. Cai, L. Chen, and X. Li, "Free market of crowdsourcing: Incentive mechanism design for mobile sensing," *IEEE Trans. on Parallel and Distributed Systems*, 2014.
- [29] M. Babaioff, N. Immorlica, D. Kempe, and R. Kleinberg, "Online auctions and generalized secretary problems," *ACM SIGecom Exchanges*, vol. 7, no. 2, pp. 1–11, 2008.

## APPENDIX

### A. Proof of Lemma 7

Since the numbers of tasks that users can complete and the respective unit costs are i.i.d., each user can be selected in the set  $\mathcal{Z}$  with the same probability. The sample set  $\mathcal{S}'$  is a random subset of  $\mathcal{U}$  since all users arrive in a random order. Therefore, the number of users from  $\mathcal{Z}$  in the sample set  $\mathcal{S}'$  follows a hypergeometric distribution  $H(n/2, |\mathcal{Z}|, n)$ . Thus, we have  $\mathbb{E}[|\mathcal{Z}_1|] = \mathbb{E}[|\mathcal{Z}_2|] = |\mathcal{Z}|/2$ . The number of tasks allocated to each user can be seen as an i.i.d. random variable, and because  $f(\mathcal{X})$  is a linear function, it can be derived that:  $\mathbb{E}[f(\mathcal{Z}_1)] = \mathbb{E}[f(\mathcal{Z}_2)] = \mathbb{E}[f(\mathcal{Z})]/2$ . The expected total payments to the users from both  $\mathcal{Z}_1$  and  $\mathcal{Z}_2$  are  $B$ . Since  $f(\mathcal{Z}'_1)$  is computed under budget  $B$ , it can be derived that:  $\mathbb{E}[f(\mathcal{Z}'_1)] = \mathbb{E}[f(\mathcal{Z}_1)] = \mathbb{E}[f(\mathcal{Z})]/2$ , and  $\mathbb{E}[p'_1] = \mathbb{E}[p]$ .

We consider two cases according to the total payment to the selected users at the last stage as follows.

**Case 1):** The budget  $B$  is exhausted.

In this case, since the price per task paid to each selected user is  $p'_1$ , so we have that

$$f(\mathcal{Z}'_2) = \frac{B}{p'_1} \geq f(\mathcal{Z}'_1) \geq \frac{\delta L}{4} \geq \frac{L}{2},$$

where the first inequality follows from the fact that the total payment to  $\mathcal{Z}'_1$  should be no more than the budget  $B$ , i.e.,  $p'_1 f(\mathcal{Z}'_1) \leq B$ .

**Case 2):** The budget  $B$  is not exhausted.

Since  $\mathbb{E}[p'_1] = \mathbb{E}[p]$ , each user  $i \in \mathcal{Z}_2$  will be allocated in  $\mathcal{Z}'_2$ . It implies that:

$$\mathbb{E}[f(\mathcal{Z}'_2)] = \mathbb{E}[f(\mathcal{Z}_2)] \geq \frac{L}{2}.$$

Combining case 1) and case 2), we have that at least  $L/2$  tasks will be allocated at the last stage in expectation.

### B. Proof of Lemma 9

We consider two cases according to the total payment to the selected users at the last stage as follows.

**Case 1):** The budget  $B$  is exhausted.

In this case, since the price per task paid to each selected user is  $p'_1$ , so we have that

$$f(\mathcal{Z}'_2) = \frac{B}{p'_1} \geq f(\mathcal{Z}'_1) \geq \frac{\delta L}{4} \geq 2L,$$

where the first inequality follows from the fact that the total payment to  $\mathcal{Z}'_1$  should be no more than the budget  $B$ , i.e.,  $p'_1 f(\mathcal{Z}'_1) \leq B$ .

**Case 2):** The budget  $B$  is not exhausted.

Because  $p'_1$  is computed over a smaller subset compared with  $p$ , we have  $p'_1 \geq p$ . Therefore, for each user  $i \in \mathcal{Z}_2$  it follows that  $b_i \leq p \leq p'_1$ , and all users in  $\mathcal{Z}_2$  will be allocated. It implies that:

$$f(\mathcal{Z}'_2) \geq f(\mathcal{Z}_2) \geq \frac{f(\mathcal{Z})}{4} \geq \frac{f(\mathcal{Z}'_1)}{4} \geq \frac{L}{2},$$

where the second inequality follows from Corollary 1, and the third inequality follows from the fact that  $f(\mathcal{Z})$  is computed over the whole user set  $\mathcal{U}$ , while  $f(\mathcal{Z}'_1)$  is computed over a smaller sample set  $\mathcal{S}'$  with the same budget  $B$ .

Combining case 1) and case 2), we have that at least  $L/2$  tasks will be allocated at the last stage with a constant probability.

### C. Proof of Lemma 10

Consider a user  $i$  with true type  $\theta_i = (a_i, d_i, \tau_i, c_i)$ , and reported strategy type  $\hat{\theta}_i = (\hat{a}_i, \hat{d}_i, \tau_i, b_i)$ . According to the *Hetero-OMG* mechanism, at each time step  $t \in [\hat{a}_i, \hat{d}_i]$ , there may be a new decision on how many tasks allocated to user  $i$ , and at what price per task. For convenience, let  $T'_t$ ,  $L'_t$ ,  $b_t^*$ , and  $S_t$  denote the end time of the current stage, the residual number of tasks, the current bid threshold, and the set of selected users respectively at time step  $t$  and before making decision on user  $i$ . Let  $\theta_{-i}$  denote the strategy types of all users excluding  $\hat{\theta}_i$ . We first prove the following two propositions.

**Proposition (a):** at some time step  $t \in [\hat{a}_i, \hat{d}_i]$ , fix  $b_t^*$  and  $L'_t$ , reporting the true cost is a dominant strategy for user  $i$ . It can be easily proved since the decision at time step  $t$  is bid-independent.

**Proposition (b):** fix  $b_i$  and  $\hat{\theta}_{-i}$ , reporting the true arrival/departure time is a dominant strategy for user  $i$ . It's because that user  $i$  is always given a payment equal to the maximum payment attained during its reported arrival-departure interval. Assume that user  $i$  can obtain the maximum payment at time step  $t \in [\hat{a}_i, \hat{d}_i]$ . Then reporting an earlier arrival time or a later departure time than  $t$  does not affect the payment of user  $i$ . However, if user  $i$  reports a later arrival time or an earlier departure time than  $t$ , then it will obtain a lower payment.

Based on the proposition (b), it is sufficient to prove this lemma by adding a third proposition:

*Proposition (c):* fix  $[a_i, d_i]$  and  $\hat{\theta}_{-i}$ , reporting the true cost is a dominant strategy for user  $i$ . According to the proposition (a), reporting a false cost at time step  $t$  cannot improve user  $i$ 's payment at the current time. Thus, it only needs to prove that reporting a false cost at time step  $t \in [a_i, d_i]$  still cannot improve user  $i$ 's payment at time step  $t'$  ( $t < t' \leq d_i$ ). In the following, we consider two cases according to whether user  $i$  is selected as a winner by reporting its true type at time step  $t = a_i$ .

**Case 1):** User  $i$  is a winner at time step  $t = a_i$ .

In this case it satisfies  $c_i \leq b_t^*$  and  $\sum_{j \in \mathcal{S}_t} f_j < L'_t$ , and it can obtain the payment  $\min\{\tau_i, L'_t - \sum_{j \in \mathcal{S}_t} f_j\}b_t^*$ . At time  $t'$  ( $t < t' < T'_t$ ), user  $i$  will obtain the payment  $\min\{\tau_i, L'_{t'} - \sum_{j \in \mathcal{S}_{t'}} f_j\}b_t^*$  if it satisfies  $b_i \leq b_t^*$  and  $\sum_{j \in \mathcal{S}_{t'}} f_j < L'_{t'}$ , otherwise it will obtain the payment 0. Because  $L'_{t'} - \sum_{j \in \mathcal{S}_{t'}} f_j \leq L'_t - \sum_{j \in \mathcal{S}_t} f_j$ , user  $i$  cannot obtain higher payment at time step  $t'$  than that at  $t$ . It implies that a user cannot improve its payment by reporting a false cost if its arrival-departure interval does not span more than one stage.

Now we consider user  $i$ 's payment at time step  $t'$  ( $T'_t \leq t' \leq d_i$ ) if its arrival-departure interval spans multiple stages. According to the proposition (a), user  $i$ 's payment at time step  $t'$  depends on  $b_{t'}^*$  and  $L'_{t'}$ . Because  $b_{t'}^*$  is independent with  $b_i$ , it only needs to consider the effect of  $b_i$  on  $L'_{t'}$ . If user  $i$  reports a false cost  $b_i$  that still satisfies  $b_i \leq b_t^*$  and  $\sum_{j \in \mathcal{S}_t} f_j < L'_t$ , then it is still accepted with payment  $\min\{\tau_i, L'_t - \sum_{j \in \mathcal{S}_t} f_j\}b_t^*$  at time step  $t$ , and thus  $L'_{t'}$  remains unchanged. If user  $i$  reports a larger bid  $b_i > c_i$  and  $b_i > b_t^*$ , then it will not be selected at time step  $t$ . In this case, more tasks will be allocated to other users, and  $L'_{t'}$  will be diminished. Therefore, user  $i$  cannot obtain higher payment at time step  $t'$ .

**Case 2):** User  $i$  is not a winner at time step  $t = a_i$ .

In this case it satisfies  $c_i > b_t^*$ , or  $\sum_{j \in \mathcal{S}_t} f_j = L'_t$ . In case  $c_i > b_t^*$ , if user  $i$  reports a false cost  $b_i$  which still satisfies  $b_i > b_t^*$ , then the outcome remains unchanged. If user  $i$  reports a lower bid  $b_i < c_i$  and  $b_i \leq b_t^*$ , then it will be accepted at price  $b_t^*$  at time step  $t$ . In such case, however, its utility will be negative. In addition,  $L'_{t'}$  remains unchanged, and thus user  $i$ 's payment at time step  $t' > t$  is not affected. In case  $\sum_{j \in \mathcal{S}_t} f_j = L'_t$ , reporting a false cost does not affect the outcome at time step  $t$  or the residual budget  $L'_{t'}$  at time step  $t' > t$ . To sum up, reporting a false cost cannot improve user  $i$ 's payment at time step  $t' > t$ .

Therefore, the above three propositions together complete the proof.